MULTIMEDIA UNIVERSITY®

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 2, 2018/2019

HPB2031 –BIOINFORMATICS ALGORITHMS I

(All sections / Groups)

6 March 2019
2:30 -4:30 PM
(2 hours)

INSTRUCTIONS TO STUDENTS

1. This question paper consists of 4 pages, including this cover page.
2. You are required to attempt all questions. All questions carry equal marks (10).
3. Write all your answers in the Answer Booklet provided.
4. You may use a calculator.

Question 1

a) Compare the filling in and traceback of Needleman-Wunsch and Smith-Waterman alignment algorithms. What should be considered when choosing the appropriate algorithm for pairwise sequence alignment? [3 marks]

b) Both BLAST and FASTA alignment algorithms derived from the Local Alignment. How does the modification in FASTA algorithm different to BLAST? Explain. [2 marks]

c) Describe the numbers represented in PAM and BLOSUM. For example, PAM100 and BLOSUM65. [2 marks]

d) Given the DNA sequences S1, S2 and S3 and their alignments as below:
S1 = TGCG
S2 = AGCTG
S3 = AGCG

m_1	m_2	m_3	m_4	m_5	m_6
T	-	G	C	-	G
-	A	G	C	T	G
-	A	G	C	-	G

m_1	m_2	m_3	m_4	m_5
T	G	C	-	G
A	G	C	T	G
A	G	C	-	G

and given the following scoring scheme

s(x,y) = 1 when x = y
s(x,y) = -1 when x != y
s(-,-) = 0: when gap vs gap
s(x,-) = -2 when x vs gap
s(-,y) = -2 when gap vs y

Based on the scoring scheme, calculate the SP score for each of the alignments shown above, and determine which alignment is better. [3 marks]

Continued .....

Question 2

a)

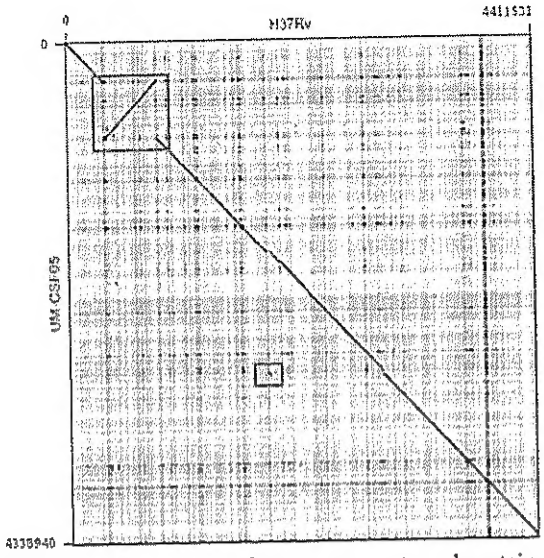

Figure 1: Figure adapted from "Chromosomal rearrangements and protein globularity changes in *Mycobacterium tuberculosis* isolates from cerebrospinal fluid", Saw et al., 2016.

What are the genomic events that can be observed from the Figure 1? Explain the meaning of each event. [3 marks]

b) Why the Illumina's paired-end sequencing technology can be used to scaffold contigs? [1 mark]

c) Discuss any TWO differences of Illumina and Pac Bio sequencing technologies. [2 marks]

d) What is the advantage of de novo assembly over reference based assembly? [1 mark]

e) List any two genome-wide alignment tools based on collinear alignment. [1 mark]

f) List any two algorithms in genome assembly? [1 mark]

g) Which parameter is important in determining the quality of genome assembly? [1 mark]

Continued .....

Question 3

a)

Table 1

	A	B	C	D	E
A	0				
B	20	0			
C	60	50	0		
D	100	90	40	0	
E	90	80	50	30	0

Table 1 illustrates the pairwise distances among five species, namely A, B, C, D and E. Reconstruct the phylogenetic tree to show the relationships among the species by using UPGMA method. Show step by step calculations and indicate the distances of the species on the branches appropriately. [5 marks]

b) List any two methods to increase reliability of phylogenetic tree inference. [1 mark]

c) What is the function to include outgroup in phylogenetic tree reconstruction? What is the selection criteria in choosing an outgroup? [2 marks]

d) Describe the methodology of bootstrapping in phylogenetic inference. Support your description on methodology with illustration. [2 marks]

Question 4

a) What are the six evidence channels in STRING database. Explain each of the evidence channel. [3 marks]

b) What are the two types of protein interactions? Explain. [2 marks]

c) How is the study on gene expression and phylogenetic profiling can be used to detect proteins interaction? [2 marks]

d) What are the three tiers in PGDB? What are the differences? [3 marks]

END OF PAPER